WORLD COMET RESEARCH GROUP

International Journal of Applied Data Science & Modern Computing
Vol. 4, No. 1, 2021
Doi: 10.67228/30713498/IJADSMC-2021PI9R8T
PP. 01-14

Original Article

Privacy-Preserving Data Mining Techniques for Sensitive Datasets

B. Sithik Shah,

Research scholar, Department of IT, PSG College of Arts and Science, Coimbatore, Tamil Nadu, India.

Received: 04-03-2021 Revised: 04-24-2021 Accepted: 05-02-2021 Published: 05-05-2021

ABSTRACT

The appraisal of Privacy-Preserving Data Mining (PPDM) has become a crucial research area in current times as a result of the incredible increase in the applications of data-driven applications that use sensitive data (health records, financial transactions, social networks, and governmental databases). Although the data mining techniques have been offering effective tools in the extraction of valuable knowledge, they facilitate great risks to personal privacy when they are applied to sensitive data. Unauthorized disclosure, inference attack, and breach of data has brought up serious ethical, legal, and regulatory issues. As a result, it is difficult to find the compromise between data utility and privacy protection. This essay outlines an extensive analysis of privacy ensuring data mining methods that allow secure privacy of sensitive data without compromising on the analysis accuracy. The paper systematically investigates the ways of anonymization, perturbation, cryptography, and hybrid privacy models. Besides that, newer privacy models include differential privacy, federated learning, and secure multi-party computation are discussed. A systematic approach is given to assess PPDM methods using privacy strength, data utility, computational complexity and scalability. The paper also reports on the findings of the experiments by comparing them, thus showing trade-offs between privacy and performance. The problems, problems under open research and direction are also discovered. The results highlight the lack of universal best practices because no single method is universally the best and the use of PPDM methods should be applied based on the application. The paper is intended to be a reference book of researchers and practitioners looking to have strong privacy preservation solutions such in sensitive data mining tasks.

KEYWORDS

Privacy-Preserving Data Mining, Sensitive Datasets, Differential Privacy, Secure Multi-Party Computation, Data Anonymization, Federated Learning.

ISSN: 3071-3498
PP: 01-14.

1.INTRODUCTION

1.1. Background

The digital data is rapidly and exponentially growing which has profoundly altered the ways of functioning of the modern industries, scientific research and the activities of society as a whole due to the proliferation of the data mining and machine learning methods. In different sectors, organizations have been relying on the use of data within decision-making as a way of attaining competitive advantages, achieving optimal levels of operational effectiveness, and providing the personal touch. The large-scale datasets have made it possible to learn more and have more precise predictive models, which has resulted in innovation in the field of healthcare, finance, e-commerce, and the smart systems. Nevertheless, much of the information gathered and processed today contains sensitive information about people, including medical records, financial transactions, biometrical identifiers, trace accesses, and behavioral patterns, and the prevalent issue with the concept of privacy and security. Using data mining methods in such sensitive data creates high exposure to privacy risks especially where such data is shared, published, or analyzed by more than one party. In situations where information of the individual like names or identification numbers have been removed, individuals can still be re-identified by pre-eminent attacks on inference and linkages, which utilize the quasi-identifiers and external data points. Such hazards are further enhanced by the rising supply of assistive information and the expandent advancement of the tools of analysis. The well-known cases of data theft, unauthorized information sharing, and personal information misuse have increased the apprehension of people and negatively affected the belief in information-driven systems. To this end, governments and regulatory authorities have come up with strict data protection act, including the General Data Protection Regulation, and the Health Insurance Portability and Accountability Act in order to promote some form of responsibility and personal privacy protection. These trends demonstrate the urgency of efficient privacy-saving data mining methods, which could allow conducting useful data analysis and maintaining strong security of the sensitive information.

1.2. Importance of Privacy-Preserving Data Mining Techniques

Privacy-Preserving Data Mining (PPDM) methods are vital towards the achievement of safe and acceptable data processing in a setting where confidential data is at stake. With the ever-growing scale of the data-driven technologies, the matter of preserving individual privacy and aidfully harvesting the knowledge became particularly relevant.

1.2.1. Protection of Sensitive Information

PPDM methods are necessary to ensure that organizational and personal information that is sensitive does not leak. Using mechanism, including anonymization and perturbation, encryption, and differentiation privacy, PPDM will make sure that data analysis does not reveal confidential attributes, including medical records and financial information, and personal identifiers. This protection prevents identity theft, discrimination and abuse of personal information.

ISSN: 3071-3498
PP: 01-14.

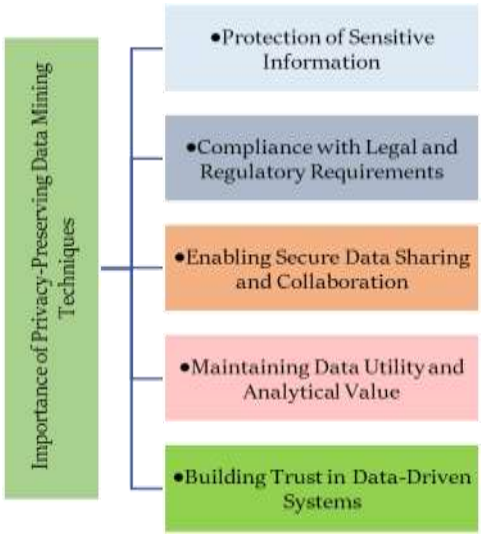

Fig 1 - Importance of Privacy-Preserving Data Mining Techniques

1.2.2. Compliance with Legal and Regulatory Requirements

Privacy maintenance has become a legal requirement due to the increased focus on data protection policies. Legal regulations and mandates like GDPR and HIPAA have a high threshold on data collection, processing, and sharing. PPDM methods help organisations become compliant with regulations by incorporating privacy protection into their data mining activities. This ensures less liability and a higher level of responsibility in the practice of data handling.

1.2.3. Enabling Secure Data Sharing and Collaboration

PPDM supports the exchange of secure data within the multiple parties which include organizations, research institutes and the distributed systems. Cryptographic techniques and federated learning among other techniques enable cross-examining data without exposing raw data. Such an ability is especially relevant to such areas as healthcare and finance where information sharing may produce better results but data sharing is often constrained by the issues of privacy.

1.2.4. Maintaining Data Utility and Analytical Value

One of the goals of PPDM is to trade privacy protection and data utility. Ideal methods of PPDM work towards conserving the statistical and structural properties of the original data whereby some valuable insights and patterns can be drawn. This trade-off allows organizations to take advantage of sensitive data in the decision-making process and reduce the effect on analysis accurateness.

1.2.5. Building Trust in Data-Driven Systems

The PPDM allows improving user trust and confidence in data-driven systems by incorporating privacy-sensitive features into the data mining processes. When people are convinced that their information is safe and is not utilized carelessly, they tend to engage in data gathering. This

ISSN: 3071-3498
PP: 01-14.

faith plays a major role in ensuring that data mining as well as machine learning technologies are adopted ethically and sustainably in the long-term.

1.3. Data Mining Techniques for Sensitive Datasets

Methods of data mining through sensitive data sets should be well developed to give significant information and reduce chances of data breach. Sensitive datasets are common in areas like the health sector, financial sector, social networks and government services where information is often characterized by personal identifiers, confidential features and behavioral facts. The oldest techniques of data mining, which include classification, clustering, association rule mining and regression, are strong analytical tools, but when used directly on sensitive data, they may accidentally reveal personal information through the output of model, some intermediate calculations, or inference attacks. Subsequently, these methods have to be adapted in specific ways to warrant privacy in the process of data mining. Privacy-conscious classification models are in need of predictive models that do not expose a individual record, frequently based on anonymized, perturbed or encrypted data. In a similar fashion, the algorithms of clustering sensitive datasets are developed to cluster the points of the data set without revealing the identity of the particular people in the clusters. The association rule mining method that is used to identify relations between attributes is particularly challenging to privacy because the identified rules can disclose sensitive relationships. Privacy-aware variations curtail the revelation of delicate trends; by managing rule generation or through noise-based techniques. Moreover, regression and statistical analysis models are modified to be privacy-aware, such that the aggregate values would not reveal any details of individuals. More complex solutions like differential privacy and cryptographic means have also contributed to the fact that the results of data mining could be applied to sensitive datasets. Differential privacy brings controlled uncertainty to the results of analytical procedures permitting productive patterns to be discovered as well as offering powerful privacy guarantees. Cryptography, such as secure multi-party computation can allow collaborative mining among multiple data owners without actually revealing raw data. Federated learning, a more recent decentralized data mining method enables models to be trained on sensitive data locally, and only aggregate updates are shared. These methods combined can help organizations to value sensitive data through accountable harnessing to serve data-driven innovation without jeopardizing privacy and security or breaking regulations.

2. LITERATURE SURVEY

2.1. Data Anonymization Techniques

One of the principle approaches to privacy protection in data publishing and mining is data anonymization. The main goal of anonymization methods is to ensure that somebody cannot recognize people again by altering or obfuscating the recognizing features in a dataset. Such techniques as k-anonymity imply that a record cannot be distinguished among at least some specific number of other records, which minimizes the possibility of identity revelation. Privacy features such as l-diversity go even more privacy enhancing by necessitating adequate diversity among sensitive

ISSN: 3071-3498
PP: 01-14.

features in anonymized groups, which aids in reducing attacks of attribute disclosure. Likewise, t-closeness makes sure that distribution of sensitive attributes in a set is not too far away from distribution of the whole dataset, and this ensures that the information the opponent can learn in terms of sensitive variables is minimized. Although difficult to implement in practice, anonymization methods are susceptible to background knowledge attacks and linkage attacks despite their simple concepts and universal usage. Moreover, active generalization or suppression usually causes significant data utility loss, which decreases the efficiency of the next data analysis.

2.2. Perturbation-Based Methods

Privacy preservation by perturbation is an approach to privacy protection that takes direct action on the data values, changing their values or altering them to stop privacy risk. These techniques make the data set uncertain with the use of random noise addition, random swapping between records, or micro-aggregation of similar data points. Perturbation techniques aim to compromise privacy and analytical utility by changing the values of individual people with global statistical characteristics that are constant. Their low calculation intensity has made them applicable to large-scale data and in real time applications. But the perversity that perturbation imparts may have adverse consequences on the quality of the data and in the end result of the mining outcomes and predictive models. Accordingly, such techniques usually cannot find a good trade-off between utility and privacy, particularly when high precision is needed in tasks.

2.3. Cryptographic Approaches

The way cryptographic methods of privacy preserving data mining endeavor to pursue their privacy preservation objectives is by ensuring that they are able to conduct the computation using sensitive data in a secure manner without going through the raw data obtained. In order to maintain a private dataset of their own, the techniques like the Secure Multi-Party Computation enable multiple owners of data to jointly analyze data or train a model without revealing their datasets. These approaches are powerful privacy assurances because they do not rely on data manipulation as their approach. Consequently, they are especially appropriate in the situations that are characterized by mutually mistrustful sides, like inter-organizational data cooperation. Nevertheless, the complexity of cryptographic methods and huge communication needs interfere with the practical implementation of cryptographic mechanisms. The constraints may cause issues with scalability and cryptography solutions are not as viable when dealing with large datasets or time-trusted applications.

2.4. Differential Privacy

Differential privacy has come out as a powerful and mathematically sound privacy protection framework in analysis of data. Rather than changing the data set, differential privacy centers around restricting the quantity of data that may be inferred about any particular record that has been computed in the result of the computation. This is normally done by introducing randomly calibrated noise in query response or learning programs. The biggest strength of differential privacy is that it

ISSN: 3071-3498
PP: 01-14.

has robust theoretical guarantee, the one that remains irrespective of the background knowability of an adversary. This has led to its mass adoption in both research studies and in industry. However, depending on the selection of priorities of privacy, the success of the protection of privacy is strong, as on the side of the opposite extreme, the accuracy of the results can be reduced because of the huge noise, and on the other extreme, the protection of privacy can be weakened due to the inadequate amount of noise.

2.5. Federated Learning and Hybrid Models

Federated learning proposes a decentralized model training paradigm where data is stored locally on participating devices/organizations. Raw data is not shared, instead, model updates are shared which would go a long way in lowering data leakage chances. This methodology is especially useful in disseminated platforms like healthcare systems, financial institutions and Internet of Things networks. Hybrid models have also been suggested to provide even greater protection to the privacy of user data by integrating federated learning with other privacy protection methods like data anonymization, encryption, and differential privacy. These combined strategies seek to get over the weaknesses of isolated strategies by using their strengths together in a complementary manner. Although hybrid models provide more privacy and security, they also impose more complexities and extra computational loads to the system and must be carefully designed to be efficient and scaled.

3. METHODOLOGY

3.1. System Architecture

The suggested methodology is based on a stacked Privacy-Preserving Data Mining (PPDM) architecture in which each of the layers is intended to address a particular phase of data processing, privacy protection and analysis. This modular architecture will help scale up, increase the security and maintenance of the system and provide a high degree of privacy.

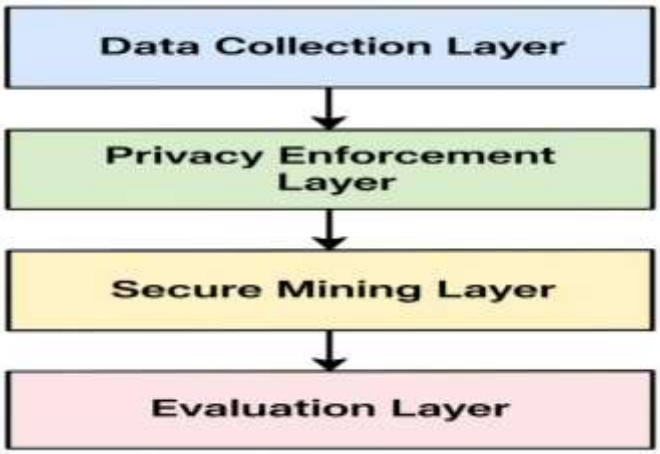

Fig 2 - System Architecture

ISSN: 3071-3498
PP: 01-14.

3.1.1. Data Collection Layer

The Data Collection Layer is meant to obtain raw information of various distributed sources including databases, sensors, or user devices. This interface is reliable in terms of data ingestion and rudimentary validation and data integrity. There is no privacy sensitive conversion done at this point so the system can acquire the correct and full information prior to including the enforcing devices.

3.1.2. Privacy Enforcement Layer

Privacy Enforcement Layer involves privacy preservation mechanisms which ensure confidentiality of sensitive information before data mining operation is carried out. This layer implements techniques like anonymization, perturbation, encryption, or differential privacy depending on a set of privacy policies. At this stage, the danger of disclosure of identity and attributes is majorly lowered by altering or securing the data.

3.1.3. Secure Mining Layer

The Secure Mining Layer analyses and gathers knowledge on privacy-protected information. Privacy preserving algorithms are used in order to make sure that the information that is of meaningful patterns and insights can be derived without revealing any confidential information. This layer facilitates approaches including classification, clustering or association rule mining without going against privacy constraints.

3.1.4. Evaluation Layer

The Evaluation Layer considers the performance of the suggested PPDM framework through the measurement of privacy protection and data utility. Information loss, accuracy, and computational efficiency are some of the metrics used to measure the performance of the system. This layer is used to give feedback in search of the best privacy choice and mining algorithms to produce a compromise trade-off between privacy and analytical utility.

3.2. Privacy Model Selection

Choosing the right privacy model can be regarded as a crucial aspect of the suggested PPDM framework. The privacy models are selected by critically managing the nature of data, regulatory limitations, and results of the intended analysis to make sure that privacy is adequately secured and does not jeopardize the utility of the data.

Fig 3 - Privacy Model Selection

ISSN: 3071-3498
PP: 01-14.

3.2.1. Sensitivity of Attributes

Data attributes sensitivity is important to define the degree of the privacy and the privacy protection type needed. More sensitive attributes like personal identifiers, medical records, or financial data are more sensitive and will require more privacy models such as: differentiated privacy or cryptographic privacy measures. The protection of less sensitive attributes can be performed through anonymization or perturbation technique, which gives a greater opportunity to conduct data analysis.

3.2.2. Regulatory Requirements

Privacy model choice is greatly affected by regulatory and legal demands. Laws and standards of data protection provide particular privacy protection to support it and avoid unauthorized disclosure. The privacy model is chosen in order to comply with these norms, where the system follows the legal requirements as well as making transparency and accountability in data management.

3.2.3. Analytical Objectives

The acceptable trade-off between the privacy and the utility of data is made with analytical objectives. Privacy models that do not alter statistics of the data can be more useful in tasks needing high precision and detailed information, whereas more invasive privacy restrictions can be accommodated by exploratory or aggregate analysis. Through establishing privacy controls and analytical objectives, the system can attain an effective and significant data mining despite privacy considerations.

3.3. Privacy-Preserving Algorithms

This section explains the privacy-sensitive algorithms that are used in the proposed framework to guarantee that data analysis is conducted in a secure way as it would incur minimum risks of sensitive information leaking.

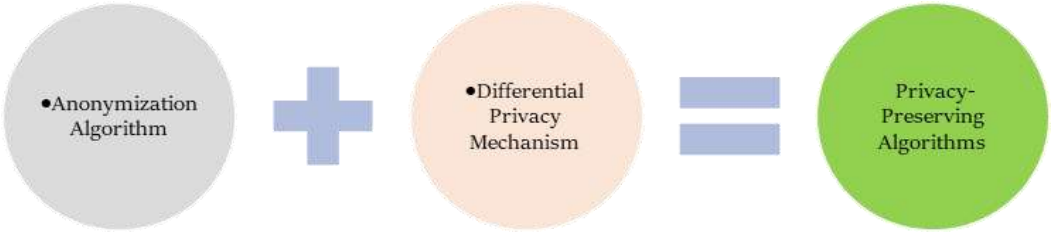

Fig 4 - Privacy-Preserving Algorithms

3.3.1. Anonymization Algorithm

The algorithm of anonymization is aimed at safeguarding quasi identifiers by means of generalization and suppression techniques to reach the k-anonymity. Generalization substitutes individual values of attributes with generalized values and suppression removes or masks values of attributes which generalization cannot adequately account. Since every record is identical to at least a

ISSN: 3071-3498
PP: 01-14.

predetermined number of such records the algorithm minimizes the possibility of re-identification. This enables protection of privacy and the data has enough utility to carry out some analytic steps.

3.3.2. Differential Privacy Mechanism

The privacy of the differential privacy mechanism is that, the controlled randomness of the output of data queries or analysis functions is introduced. Rather than exposing the real output of a function calculated on the data, random noise generated using a Laplace distribution is introduced to the actual output of the function. The level of noise is governed by the sensitivity of the function which is the largest resultant change in the output with the variation of a single record and the privacy budget which determines the strength of privacy protection. This mechanism guarantees the negligence of any individual record in the final output to give a high sense of privacy.

3.4. Evaluation Metrics

The effectiveness of the suggested privacy-ensuring data mining structure is measured on the basis of numerous measures that all quantify the strength of privacy, the efficiency of its analysis, and its efficiency in terms of computation.

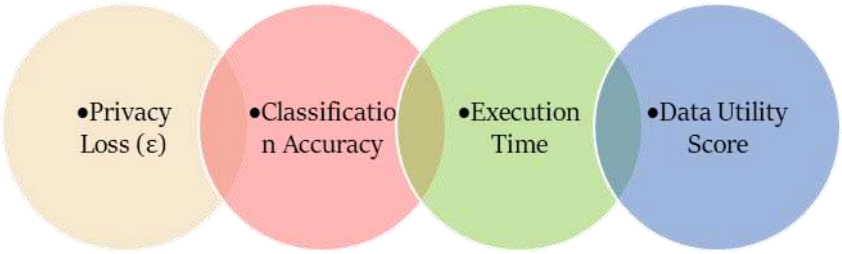

Fig 5 - Evaluation Metrics

3.4.1. Privacy Loss (ε)

The loss of privacy is a state of privacy maintenance by the system, especially under differential privacy systems. The lesser the value of privacy loss, the greater the privacy guarantees, as the former restricts the impact of an individual record on the result. This measure will be used to determine the trade-off between privacy and data accuracy to ensure whether the chosen privacy parameters are adequate to address the requirements of the desired security level.

3.4.2. Classification Accuracy

The measures of classification accuracy are used to evaluate the efficiency of data mining algorithms on privacy protected data. It represents the ratio of instances that are correctly classified with the total number of instances. This measure is required to assess the compatibility of preserving the predictive power of the model following the transformation of the data by privacy-saving methods.

ISSN: 3071-3498
PP: 01-14.

3.4.3. Execution Time

Execution time, assesses the efficiency of the proposed framework by computing the overall time it takes to implement privacy mechanisms as well as conduct data mining processes. Reduction of the time of execution implies enhanced scalability and applicability to large datasets or real-time use. This indicator is especially crucial in making comparisons on the overhead added by various privacy-saving methods.

3.4.4. Data Utility Score

The utility score data is used to measure the usefulness of the transformed data at the post privacy enforcement stage. It is a measure of the similarity of the privacy-protected data to the original dataset in terms of statistics and structure. When the score of data utility is high, this means that there is more efficient balance in privacy protection and the retention of useful information to be analyzed with data.

4. RESULT AND DISCUSSION

4.1. Comparative Analysis

It is in this section that a comparative analysis amongst different Privacy-Preserving Data Mining (PPDM) methods is drawn on the basis of the degree of privacy, the usefulness of the data, and their cost in terms of percentage to show how they perform accordingly.

Table 1: Comparative Analysis

Technique	Privacy Level (%)	Data Utility (%)	Computational Cost (%)
k-Anonymity	60%	65%	30%
Differential Privacy	85%	65%	55%
SMPC	95%	85%	90%
Federated Learning	85%	85%	60%

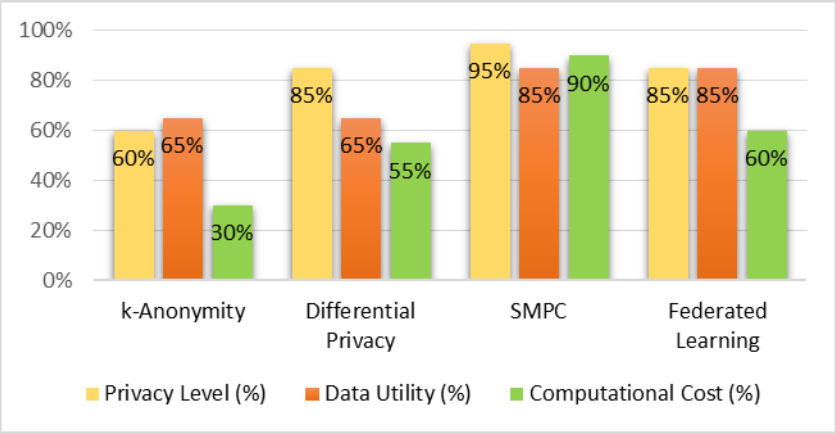

Fig 6 - Comparative Analysis

ISSN: 3071-3498
PP: 01-14.

4.1.1. k-Anonymity

The level of privacy protection offered by k-Anonymity is moderate because it guarantees that a given record cannot be distinguished against any other record of similar type. It provides fair protection against direct identification as its privacy level of 60 is quite high but still is susceptible to background knowledge and attribute disclosure attacks. The utility in data of 65% shows that anonymized data still has a reasonable degree of use in analysis since overall trends are still promoted by generalization and suppression. K-anonymity is applicable in large volumes of data and real time applications that require speed due to its low calculation cost of 30%.

4.1.2. Differential Privacy

The level of privacy of a differential privacy is 85%; this is because through controlled noise injection the effect of a single record on the end result of analysis is restricted. This high privacy assurance is even as against persons who may be adversaries but with prior knowledge. The utility of the data is moderate of 65, since the noise added has the potential of slightly reducing accuracy, especially in fine-grained analysis. The computational cost of 55 percent, coupled with the difference between privacy and practical efficiency, has made differential privacy to introduce moderate overhead.

4.1.3. Secure Multi-Party Computation (SMPC)

SMPC runs on a very high privacy of 95 per cent since it can be used to conduct collaborative computation without exposing raw data to parties involved. Such cryptographic power is that sensitive data can be analyzed with confidence without any loss of confidentiality. The utility of data is high with 85 percent because operations are carried out on the real data as opposed to disturbed data. Nonetheless, such a high level of privacy is at the cost of a high computational cost of 90% as a result of heavy cryptographic workloads and a large communication overhead that could be restraining scaling.

4.1.4. Federated Learning

Using Federated learning for learning, the privacy level of 85 percent means that information is localized and presented by updated model rather than unfiltered information. The model has a strong capability to mitigate data exposure risks and help in collaborative model training. Having a data utility 85, federated learning has good analytical behaviour, with models being trained on local original data. It is a moderate computational cost of 60 probability that represents overhead of model update at every iteration and model-communication, which is an even-balanced distributed and privacy sensitive environment.

4.2. Discussion of Findings

As revealed by the experiments, there exists a significant trade-off between the privacy-preserving methods of data mining methods evaluated, as not one of the methods is an ideal answer to all aspects of privacy, utility and efficiency. Differential privacy proves to have a good theoretical

ISSN: 3071-3498
PP: 01-14.

privacy guarantee whereby the addition or absence of any individual record to the analysis has a small difference on the results. This feature renders this exceptionally strong against adversaries who have some background knowledge. The results however show that the tuning of the privacy budget and its selection is very critical in determining the success of differential privacy. A very stringent budget on privacy can bring undue noise, causing the loss of utility of the data and analysis accuracy to become apparent, but a lenient budget might dilute privacy. Hence, a reasonable and balanced deployment is needed, corresponding to the sensitivity of the information and analytical intentions. The highest privacy level is the use of cryptographic methods, especially the Secure Multi-Party Computation, which makes it possible to perform the precise calculations without revealing raw data. These findings validate the conclusion that such methods ensure the utility of data because they do not cause distortion of data in the course of analysis. Even with these benefits, cryptographic protocols can have a serious effect on the performance as a result of their associated computational and communication overhead. Execution time and resource consumption scales up with the size of the dataset and parties involved to a significant extent, making these techniques highly impractical to execute in large-scale or real- time environments. The fusion of various privacy-preserving methods like federated learning, anonymization, and differential privacy are linked to hybrid models that can serve as a promising alternative. The findings reveal that hybrid strategies can effectively trade-off privacy and utility by harnessing the power of each of the component methods and limiting the weaknesses. Hybrid models offer good privacy with low computation and reduced data exposure through the sharing of computations, which reduces the computational burden, applying restricted privacy mechanisms, and a high level of privacy benefits. On the whole, the results indicate that abduktive and hybrid privacy preserving frameworks are better adapted to practical applications, where the sensitivity of the data, regulatory conditions, and performance needs should be handled at the same time.

5. CONCLUSION

The given paper provided a solid review and comparison of privacy-conscious data mining algorithms that aim to ensure the security of sensitive datasets and also allow the analysis of data to be performed. The research shedding light on the nature of fundamental trade-offs between privacy protection, data utility, and computation efficiency by methodically discussing traditional strategies of anonymization, perturbation-based methods, cryptography and current privacy models like differential privacy and federated learning. The methods of anonymization and perturbation have been demonstrated to have simplicity and lack of severe computational expenses but are susceptible to inference attacks and analytic error. Conversely, cryptographic solutions offer a high level of privacy and data fidelity, but are expensive to compute and communicate which makes them ineffective on large scale. Differential privacy provides striking theoretical guarantees though it must be calibrated to strike a balance between privacy and utility.

In this analysis, the paper has shown that the decision of how to adopt the right PPDM technique should be informed by sensitivity of the information, limitation of the regulations and the

ISSN: 3071-3498
PP: 01-14.

intended analysis goals. The evaluation framework to be used in this research undertaking introduces a systematic approach to comparing the use of PPDM techniques in terms of the main key performance indicators (KPI) to help the practitioners and researchers make effective decision-making in bringing privacy-sensitive solutions within the practical context. The methodology assists in a fair evaluation of the solution considering the security and performance issues because it considers the loss of privacy, utility of data and classification accuracy as well as the time of execution. The comparative findings also highlight the idea that hybrid methods that combine several privacy-sensitive tools in one tend to reach a more achievable trade-off in comparison to single methods, and hence in most complex and large data instances, hybrid methodology is especially appropriate.

Future studies in privacy-conscious data mining are likely to prospect on dynamic privacy models that optimally change privacy parameters depending on sensitivity of data, the needs of the users and risks in context. Another avenue of rich opportunities is the union of privacy-sensitive algorithms with deep learning, which will be relevant in the radiation of increased calls on the privacy of the analysis of high-dimensional and unstructured data. Also, explainable PPDM systems are becoming more relevant, and transparency and interpretability is necessary to establish trust in privacy- conscious analytical systems. Lastly, closer planning of PPDM methods and implementing the mechanisms of legal and ethical compliance will be essential in order to guarantee compliance with the changing data protection laws and social demands. All of these future avenues are geared toward making privacy preserving data mining systems more convenient, responsible and resolute.

REFERENCES

[1] Samarati, P., & Sweeney, L. (1998). *Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression*. Proceedings of the IEEE Symposium on Research in Security and Privacy.

[2] Sweeney, L. (2002). *k-anonymity: A model for protecting privacy*. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5), 557–570.

[3] Machanavajjhala, A., Gehrke, J., Kifer, D., & Venkitasubramaniam, M. (2007). *l-diversity: Privacy beyond k-anonymity*. ACM Transactions on Knowledge Discovery from Data, 1(1), 3.

[4] Li, N., Li, T., & Venkatasubramanian, S. (2007). *t-closeness: Privacy beyond k-anonymity and l-diversity*. Proceedings of the IEEE 23rd International Conference on Data Engineering.

[5] Fung, B. C. M., Wang, K., Chen, R., & Yu, P. S. (2010). *Privacy-preserving data publishing: A survey of recent developments*. ACM Computing Surveys, 42(4), 1–53.

[6] Agrawal, R., & Srikant, R. (2000). *Privacy-preserving data mining*. Proceedings of the ACM SIGMOD International Conference on Management of Data.

[7] Domingo-Ferrer, J., & Torra, V. (2005). *Ordinal, continuous and heterogeneous k-anonymity through microaggregation*. Data Mining and Knowledge Discovery, 11(2), 195–212.

[8] Verykios, V. S., Bertino, E., Fovino, I. N., Provenza, L. P., Saygin, Y., & Theodoridis, Y. (2004). *State-of-the-art in privacy preserving data mining*. ACM SIGMOD Record, 33(1), 50–57.

[9] Lindell, Y., & Pinkas, B. (2000). *Privacy preserving data mining*. Journal of Cryptology, 15(3), 177–206.

[10] Goldreich, O. (2004). *Foundations of cryptography: Volume 2, basic applications*. Cambridge University Press.

[11] Dwork, C. (2006). *Differential privacy*. Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP).

ISSN: 3071-3498
PP: 01-14.

[12] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). *Calibrating noise to sensitivity in private data analysis*. Proceedings of the Theory of Cryptography Conference.

[13] Dwork, C., & Roth, A. (2014). *The algorithmic foundations of differential privacy*. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407.

[14] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). *Communication-efficient learning of deep networks from decentralized data*. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS).